SYSTEM AND METHOD FOR INDEXING AND QUERYING STRUCTURED TEXT

ABSTRACT OF THE DISCLOSURE

5   An indexer indexes a database of documents, and a search engine searches the database of documents. Nesting level information stored in index entries is used to identify, and match together, start and end meta words comprising fields at assorted nesting levels within a document. Based on a query specifying words to be found within fields, spatial criteria are applied to the identified meta words to determine if the specified words are found within the specified fields. A subset of the documents have nested fields, and each nested field has an associated start meta word and end meta word. Each meta word has an associated nesting level. Each document is indexed by parsing the document to determine locations within the document of words and meta words, as well as the nesting level associated with each meta word. An index is generated that has word entries, meta word entries, and generic meta word entries. The meta word entries indicate locations within the documents of an identified meta word, as well as the nesting level of the meta word. The generic meta word entries identify locations within the document of a class of meta words, including meta words at all nesting levels of the meta words within the document. For each identified location within the generic meta word entry, the generic meta word entry also includes nesting level information associated with the meta word at the identified location.